Summer 5-2019

MACHINE-LEARNING AND STATISTICAL SUBGROUP ANALYSES OF VASOPRESSOR CHOICE FOR THE TREATMENT OF NON-TRAUMATIC SUBARACHNOID HEMORRHAGE USING ELECTRONIC HEALTH RECORDS

EMY ANU THOMAS
UTHealth School of Public Health

MACHINE-LEARNING AND STATISTICAL SUBGROUP ANALYSES OF VASOPRESSOR CHOICE FOR THE TREATMENT OF NON-TRAUMATIC SUBARACHNOID HEMORRHAGE USING ELECTRONIC HEALTH RECORDS

by

EMY ANU THOMAS, BE, ME

APPROVED:

Jose-Miguel yamal Digitally signed by Jose-Miguel yamal Date: 2019.05.01 10:07:30 -05'00'

JOSE-MIGUEL YAMAL, PHD

Xianglin Du Digitally signed by Xianglin Du Date: 2019.04.29 16:27:28 -05'00'

XIANGLIN L DU, PHD

DAVID AGUILAR, PHD

MACHINE-LEARNING AND STATISTICAL SUBGROUP ANALYSES OF VASOPRESSOR CHOICE FOR THE TREATMENT OF NON-TRAUMATIC SUBARACHNOID HEMORRHAGE USING ELECTRONIC HEALTH RECORDS

by

EMY ANU THOMAS

ME, Indian Institute of Technology Madras, 2011

BE, University of Calicut, 2008

Presented to the Faculty of The University of Texas Health Science Center

School of Public Health

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

THE UNIVERSITY OF TEXAS HEALTH SCIENCE CENTER
SCHOOL OF PUBLIC HEALTH
Houston, Texas

May, 2019

MACHINE-LEARNING AND STATISTICAL SUBGROUP ANALYSES OF VASOPRESSOR CHOICE FOR THE TREATMENT OF NON-TRAUMATIC SUBARACHNOID HEMORRHAGE USING ELECTRONIC HEALTH RECORDS

Emy Anu Thomas, BE, ME, MS

The University of Texas Health Science Center

School of Public Health, 2019

Thesis Chair: Jose-Miguel Yamal, PHD

Abstract: Subarachnoid hemorrhage (SAH) is a life-threatening stroke caused by bleeding in the subarachnoid space. Delayed cerebral ischemia, a risk factor for death during non-traumatic SAH treatment, is often prevented by inducing hemodynamic augmentation through the administration of vasopressors. The three common vasopressors administrated to SAH patients are dopamine, phenylephrine, and norepinephrine. A recent study using Electronic Health Records (EHR) of non-traumatic SAH patients identified that the administration of phenylephrine is associated with lower hospital mortality; however, this study did not consider heterogeneity in treatment effects. Therefore, the goal of this paper is to study the heterogeneity in vasopressor treatment effects by exploratory subgroup analyses. The subgroups were identified by studying the vasopressor and pretreatment covariates interaction effects. We employed a machine learning and a generalized linear model approach to study the interaction effects. We also employed propensity scoring and

inverse probability weighting to minimize the confounding and selection biases in treatment effect estimates due to inherent pretreatment group differences in the data. Our results showed that dopamine had the highest mortality among those who did not have ondansetron (a pretreatment medication used to prevent nausea) compared to phenylephrine and norepinephrine; whereas norepinephrine has the highest mortality among those who had ondansetron compared to phenylephrine and dopamine. However, for the subgroup who did not have ondansetron but had fentanyl and lidocaine, there was no significant difference in the mortality rate between vasopressor treatment group. Overall, those who had ondansetron had better survival compared to those who did not have ondansetron with respect to all the three vasopressors.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

Study Background

Stroke and SAH

Stroke is the fifth leading cause of death in 2016, and it accounts for 1 in 20 deaths across the United States (U. S.). [4] Two main types of stroke are ischemic stroke and hemorrhagic stroke. Ischemic stroke is a common stroke caused by a blood clot that blocks a blood vessel in the brain. It accounts for 80% of all cases and it transpires because of oxygen deprivation in the brain. In comparison, a hemorrhagic stroke occurs when a blood vessels ruptures, and bleeds into the intracerebral space (referred to as intracerebral hemorrhage, ICH) or subarachnoid space (referred to as subarachnoid hemorrhage, SAH). SAH constitutes 5-10% of strokes annually reported in the U. S.[1] SAH can occur due to trauma or non-trauma conditions, differing in their underlying pathophysiology. Among the non-trauma SAH cases, 85% is due to a ruptured aneurysm, while the remaining 15% are due to nonaneurysmal perimesencephalic bleeding and other causes.[3] The estimated incidence of nontraumatic SAH in the U. S. during 1979-2008 was 7.2 to 9.0 per 100,000 per year.[2] Metanalysis of 51 studies through October 2005 showed a worldwide incidence of nontraumatic SAH of approximately 9 per 100,000 person-years.[5]

A retrospective cohort study of non-traumatic SAH patients reported an in-hospital mortality rate of 14%.[2] Concurrently during SAH, patients can also experience delayed cerebral ischemia (DCI), one of the primary causes of in-hospital mortality. DCI is linked to cerebral vasospasm, a vascular constriction of cerebral vessels after SAH. Usually, between 20-40% of the SAH patients would develop cerebral vasospasm.[6] Among those affected, 50% will die or suffer permanent neurological damage.[6] However, DCI in patients presenting with SAH could often be prevented by inducing hemodynamic augmentation to maintain cerebral perfusion. Administration

of vasopressors is a widely held medical practice for inducing hypertension.[7,8] Three common vasopressors administered are dopamine, phenylephrine, and norepinephrine, but which of these vasopressors should be chosen as the first-choice treatment still remains unknown.

Currently, no prospective comparative study on vasopressor treatment choice for SAH patient are available. However, it is believed that vasopressor choice is patient specific, and it depends on the one's prior diagnosis and medication.[9] A recent retrospective study, I have been a part of, using the Cerner Health Facts Electronic Health Records (EHR) data (n=2535) found that phenylephrine is associated with lower hospital mortality (under review). However, we cannot generalize this observation to all subjects because each subject may have a different treatment response. Therefore, the goal of this paper is to study the differential effect of vasopressor by exploratory subgroup analyses.

Electronic Health Record (EHR), Cerner Health Facts

EHR is a longitudinal electronic record of patient health information generated by one or more encounters in any healthcare delivery setting.[14] Although EHR has been primarily used for structured billing and records, with advances in big data analysis and machine learning methodologies, it is now being utilized for quantitative analysis to improve patient care and to identify optimal treatments or treatment strategies, including outcomes research and epidemiological studies. A major provider for the EHR systems in the U. S. is the Cerner Corporation in North Kansas City, Missouri.[15] In this study, we used the SAH patient records from the Cerner health facts database hosted by SBMI (School of Biomedical Informatics) at UT Health Science Center. This is a publicly available relational database comprising of multicenter de-

identified 106 million health records. The data contain patient records across 700 participating Cerner client hospitals and clinics from 2000 to 2015 in the U. S.

EHR includes individual patient records on demographics, diagnosis history, medication history, clinical procedure, vital signs, lab records, and billing information. Each patient record follows the key standards for clinical vocabularies, such as Current Procedure Terminology (CPT) codes for clinical procedures, International Classification of Diseases (ICD) for diagnoses and Logical Observation Identifiers, Names and Codes (LOINC) for ordering lab tests. Since these standards are implemented uniformly across medical institutions, patient data from different institutions can be combined to make nationally representative data. Aggregated EHR data such as Cerner health facts then give access to a large amount of data that are useful for studying rare events such as SAH. EHR data also have a large number of variables that can be studied simultaneously. For issues such as missing data and data quality[16,17] in the EHR data and the presence of correlated covarites, an appropriate staitical method such lasso modeling and propensity scring were employed in our analysis.

Specific Aims

The objective of this thesis is to identify patient subgroups for which vasopressor has differential treatment outcomes. Identifying these patient subgroups would help to inform the optimum vasopressor choice among dopamine, phenylephrine, and norepinephrine from a patient-specific perspective. The differential effect of vasopressor treatment for SAH patients was studied by exploratory subgroup analyses in this paper. To conduct subgroup analyses, we utilized two approaches: a generalized linear model and a machine learning approach. In the generalized linear model approach, we employed lasso regression for feature selection; and applied propensity scoring to reduce the effects of observed confounding and to obtain unbiased estimates of the

average treatment effects. False Discovery Rate (FDR) approach[22] was then used to correct for multiple comparisons. We also employed a machine learning method, a recursive partitioning method for classification and regression. The theory behind lasso, propensity scoring, and recursive partitioning method are discussed in the Methods section.

The specific aims are as follows:

1. Conducting a statistical, retrospective exploratory subgroup analysis of vasopressor treatment effects in SAH patients on mortality outcome via a generalized linear model (GLM) approach:
    a. Constructing covariate balanced data using propensity score modeling and inverse probability treatment weighting.
    b. Identifying demographic, medication, and vital sign variables associated with patient mortality.
    c. Identifying significant interactions between different vasopressor treatments and covariates, adjusting for variables selected in 1b. Interaction p-values were adjusted for multiple testing.
    d. Studying the magnitude of treatment effect among the subgroups identified through significant interaction effects.
2. Conducting a machine learning, data-driven tree-based recursive partitioning approach to explore heterogeneity in vasopressor treatment effects.
3. Comparing the subgroups from (1) GLM and (2) machine learning approaches

We hypothesize that there is an interaction effect of pretreatment covariates on the relationship between vasopressor treatment effect and mortality. Any identified interaction effect

will show the presence of subgroups within the SAH patients. This could then generate potential new hypotheses for future studies of personalized medicine in SAH patients receiving vasopressor treatments.

Public Health Significance

SAH patients have a high mortality rate. Currently, patients are mainly treated using three vasopressors: dopamine, phenylephrine, and norepinephrine. In the absence of clinical trials on vasopressor choice, the decision has often to be made by physicians based on their experience and the patient's hemodynamic status, comorbid conditions, and institutional preferences, which may affect patient outcomes after the diagnosis of SAH. The data-driven subgroup analysis using the EHR data in this study can then highlight the differences in different patients' responses to vasopressor treatment. Results from our analysis can be used as hypothesis generating for the guideline in designing future clinical trials; and choosing the optimal vasopressor for a group of SAH patients. In the long term, our study also has the potential to improve SAH treatment and reduce the mortality rate among SAH paitents.

Method

Subgroup analysis is a topic of great interest specifically with an eye towards more personalized medicine. It is useful in determining the heterogeneity of treatment effects, and it is critical to identify the positive and negative effects of the treatment. Subgroup analyses can be conducted either as pre-specified or during post-hoc reviews. In a pre-specified subgroup analysis, hypotheses about which subgroups to examine are predefined. In clinical trials, subgroups are often defined at the design stage based on prior insights or the exploratory analyses of other datasets. In

this thesis, we will conduct an exploratory analysis for subgroup identification by examining all possible subgroups.

However, there exist several limitations and pitfalls of post-hoc subgroup analyses. A major limitation of a subgroup analysis is its risk of increasing the probability of finding statistically significant findings due to chance. Due to multiple tests, a study can run into a risk of false positives during the post hoc analysis. Nevertheless, multiplicity correction using FDR can reduce this effect. Moreover, if we test the statistical interaction between treatment and some subgroups first, and then test within the subgroups only if a certain interaction is significant, we could reduce the number of tests conducted. So, in this thesis, we used both multiplicity correction and limited number of tests to avoid false positive during the post hoc analysis.

Subgroup analysis can also lead to false negatives due to inadequate power. Brookes ST et al[10], used a simulation study to quantify the power of interaction test in trials designed to detect the overall treatment effects. They looked at different scenarios of the overall treatment and subgroup effects using both continuous and binary outcomes, and they found that if a trial has 80% power for the overall effect, the power to detect the interaction effect of the same magnitude is only 29%. They also pointed out that to detect the interaction effect of the same magnitude with the same power as the overall effect, the sample size needs to be fourfold.[10] However, even with these limitations, subgroup analyses are still often conducted to extract information from data for hypothesis generation. Of course, one has to follow several guidelines while reporting and interpreting the results.[11,12,13] Therfore, in this thesis, to identify subgroups, we evaluated multiple covariate interaction effects on vasopressor treatment choices for a SAH patient outcome model.

When constructing the SAH patient outcome model to account for confounding, it would be appropriate to incorporate as many variables as available. However, there is a risk of overfitting the model, resulting in a numerically unstable estimate. Overfitting could also reduce the power of prediction for future data due to the presence of noise[18,19] and violation of the principle of parsimony. A parsimonious model should include no more predictors than necessary. Peduzzi et al[20] proposed a guideline regarding the number of predictors that should be included in statistical models. This guideline suggests including at least 10 outcomes of each type for every independent variable to avoid over-estimated and under-estimated variance. Hosmer and Lemeshow[21] suggested the number of parameters $(p+1)$ in a logistic regression model should be $\leq \min(n_0, n_1)/10$, where, n_1 is the number of events of type 1, and n_0 is the number of events of type 2 of a binary dependent variable. In the SAH data, n_1= 1895 and n_0 = 2720; hence, by above mentioned criteria, we could include a maximum of 189 parameters. Hence it is important to choose an appropriate statistical method to find important predictors to fit a parsimonious model. Another important issue that needs to be addressed is confounding. Confounding occurs if the SAH patients who received one vasopressor differed systematically from those receiving a different vasopressor in terms of pretreatment covariates. However, this could be addressed by removing or minimizing the effect of confounding through choosing a suitable statistical method such as propensity scoring.

Overview of the proposed analysis

Generalized linear models and machine learning approaches are used to identify the significant interaction effect between vasopressor and pretreatment covariates within the SAH patient data. The generalized linear model analysis included four steps. First, the propensity score, which is

defined as the probability of receiving a treatment conditional on observed pretreatment covariates. It was calculated using Generalized Boosted Models (GBM), a flexible estimation method that can adjust for many pretreatment covariates.[23] Prior to GBM, we used the L_1-penalized multinomial logistic regression (GLM Multinomial Lasso) to identify important factors associated with vasopressor treatment. These selected variables were used as input variables in GBM to obtain the propensity score. Obtained scores were then used as weights in the mortality outcome model whicn test for interactions in the third step. The second step involved using L_1-penalized logistic models (GLM Logistic Lasso) to select the important factors associated with mortality. The selected variables were then used as the main effects for the outcome model in the third step. In the third step, a base outcome model was constructed using weights from the first step, the main effects from the second step and the vasopressor treatment variable. Further, an interaction model was also constructed with the treatment interaction with each main effect covariate identified from the second step. Finally, the base model and interaction model were compared using the Likelihood ratio test (LRT). In order to account for the multiplicity correction in LRT, the FDR adjusted p-value was used to identify significant interaction effects. Any significant interactions identified would highlight the difference in the effect of vasopressor on mortality conditional on the presence or absence of that main effect. Further, we conducted Model-based Recursive Partitioning (MOB), a machine learning approach for subgroup analysis. We then compared the generalized linear model result with the MOB result.

Data Summary

In this study, we used the SAH patient records queried from the Cerner Health Facts® database. This database is hosted by SBMI, UT Health Science Center and accessed through an appropriate IRB approval (IRB HSC-MS-18-0124). Our data included de-identified patient-level electronic

health records from 700 hospitals and clinics that use Cerner Corporation's EHR system. These correspond to 39,017 patients with at least one SAH encounter.

Inclusion/Exclusion criteria: From the 39,017 SAH patient records, our study population was identified based on the following criteria: (1) Only records with diagnosis history on ICD-9 code (430,800.2x, 800.7x, 801.2x, 801.7x, 803.2x, 803.7x, 804.2x, 804.7x, 852.x) with medication history of at least one of the vasopressor medications norepinephrine, phenylephrine or dopamine were included (N=4,850). (2) SAH patients below the age of 17 were excluded. (3) SAH patients who received multiple vasopressor treatments simultaneously as initial treatment were excluded (n=40). (4) SAH patient who had a diagnosis of the traumatic cause of SAH were excluded (n=2176).

Date cleaning: Observations with Mortality = "Unknown" (n=93), Gender = "Unknown" (n=2) and those with missing medication information were all omitted from the analysis. Further, pretreatment medication variables with less than 10 occurrences were also excluded.

Study population: The study population included non-traumatic SAH patients over the age of 17 with a single encounter or multiple encounters for SAH (N= 2415). When a patient had multiple encounters separated by less than a day, the encounters were then combined into a single encounter otherwise only data from the first encounter were used in the analysis. A subsample of size N= 940 included the data of baseline heart rate, baseline systolic and diastolic blood pressure, and all these data were used for sensitivity analysis.

Exposure: The primary exposure of interest is the first vasopressor administered to eligible patients, either phenylephrine, dopamine, or norepinephrine.

Outcome: The outcome for this study is the mortality rate which is defined as death at hospital or discharge to hospice. Mortality will be coded as 1 (death or discharge to hospice) and 0 (survived).

Our final cleaned data consisted of 2,415 patients' records, and each was identified using an individual identifier (variable *Patient_SK*). The data also included *age in years*, *race* (African American=549, White= 1609, other = 257), *marital status* (divorced=258, married= 1098, single= 637, widowed=180, unknown= 242), *gender* (female=1536 , male=879), mortality, vital signs (heart rate, systolic blood pressure and diastolic blood pressure) and total fluids administered (*sum_IVF*). Appendix I shows a summary of all the variables. Additionally, in the Cerner database, because we did not have timing information for diagnoses, we could not differentiate whether the observed diagnoses were pretreatment or post-treatment factors. Hence, diagnosis factors were not included in our analysis.

Statistical Analysis

The analysis was performed using R software (R Foundation for Statistical Computing, Vienna, Austria). Specifically, we used ANOVA or Kruskal-Wallis test to compare the continuous variables among the vasopressor treatment groups. For categorical factors, we used the Chi-square test. The following subsection describes the various statistical methods used in our analysis.

Least absolute shrinkage and selection operator (Lasso)

L1-penalized regression (also referred to as Lasso) was used to select the important factors associated with the dependent variable of interest. We applied Lasso prior to GBM and the treatment-covariate interaction studies to identify the important factors associated with vasopressor and mortality, respectively. Lasso is a regularization and feature selection method first formulated by Robert Tibshirani in 1996. In high dimensional data such as EHR where the number

of features is larger than the number of observations, a sparse statistical model such as lasso would be appropriate since it involves regularization techniques of penalizing the absolute size of the regression coefficients. Lasso shrinks the less important feature's coefficient to zero, thus removing some less important features altogether. The purpose of the shrinkage is to prevent overfit arising due to collinearity of the covariates or high-dimensionality. Lasso could minimize the sum of squared error subject to an l_1 penalty term $\lambda \sum_j^k |\beta_j|$, where $\lambda \geq 0$ is a tuning parameter. When λ =0, the Lasso estimate is equivalent to the least squares estimate. When λ is sufficiently large, coefficients are forced to be exactly equal to zero, which help us attain dimensionality reduction.

Lasso estimator for GLM

The Generalized Linear Models (GLM) is a broad class of the linear regression models that allow analysis of different types of univariate response. Standard linear regression is a special case of GLM for the normally distributed continuous response; likewise, logistic regression is GLM for binary response, and log-linear regression is GLM for counts data. GLM consists of three main components:

1. Random component: GLM assumes that the probability distribution of the response variable is from an exponential family distribution including normal, Bernoulli, binomial and Poisson distributions. The random component defines a probability distribution for the response variable, Y_i. In a logistic regression model, the probability distribution of the response variable follows a Binomial distribution.

2. Systematic component: Systematic component specifies the effect of an explanatory variable on the mean of Y, which can be expressed as a linear combination of these predictors:

$$\eta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

2. Link function: This component specifies a function g(.) that relates the expected value of Y, μ to the linear predictors. That is, a link function connects the systematic and random component as

$$g(\mu) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

For logistic regression, the link function is the logit function.

$$g(\mu) = logit(\mu) = \log\left(\frac{\mu}{1-\mu}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k$$

For GLM, lasso estimator is defined by penalizing the negative log-likelihood with the l_1-norm. The negative log-likelihood is

$$-\sum_{i=1}^{n} \log(p_\beta(Y_i|X_i)) \qquad (1)$$

The equation (1) in terms of the loss function can be expressed as

$$n^{-1} \sum_{i=1}^{n} \rho_\beta(X_i, Y_i) \qquad (2)$$

$$\rho_\beta(x, y) = -\log(p_\beta(y|x)) \qquad (3)$$

Then, the l_1-norm penalized Lasso estimator is defined as

$$\hat{\beta}(\lambda) = \underset{\beta}{argmin} \left(n^{-1} \sum_{i=1}^{n} \rho_\beta(X_i, Y_i) + \lambda \|\beta\|_1\right) \qquad (4)$$

For a logistic regression where $Y_i|X_i \sim Binomial\ (1, \mu_i)$, the negative log-likelihood equals

$$-\sum_{i=1}^{n} \log(p_\beta(Y_i|X_i)) = \sum_{i=1}^{n} \left\{-Y_i f_\beta(X_i) + log\left(1 + exp\left(f_\beta(X_i)\right)\right)\right\} \qquad (5)$$

Where $f_\beta(X_i) = logit(\mu) = \log\left(\frac{\mu}{1-\mu}\right) = \sum_{j=0}^{k} \beta_j x_j$, and the corresponding loss function for (5) is

$$\rho_\beta(x, y) = -y\left(\sum_{j=0}^{k} \beta_j x_j\right) + log\left(1 + exp\left(\sum_{j=0}^{k} \beta_j x_j\right)\right)$$

Then, the lasso estimator for the logistic model is defined as that in (4).

Propensity Score (PS)

One of the challenges in the analysis of observational data such as EHR is the presence of confounding bias. Because treatments were not given randomly, confounding can occur if the underlying distribution of pretreatment covariates differ between the vasopressor treatment groups. Hence, direct comparisons between the groups could result in a biased estimate of treatment effects.[24] However, propensity scoring (ps) methodology can be used to create comparison groups which have similar values for observed confounders. In this way, the effect of the type of treatment received could be isolated. PS was first introduced by Rosenbaum and Rubin in 1983. It is defined as the conditional probability of assignment to a treatment given a vector of observed covariates.[24] The propensity score, $e(X_i)$ for the subject i, is given by

$$e(X_i) = P(Z = k|X_i)$$

where $Z = k$ denotes kth treatment, and X_i is the covariate matrix of that subject, and it is assumed that X's and Z's are independent.[25] PS modeling also requires positivity and exchangeability assumptions. The positivity assumption states that all the subjects should have a non-zero probability of receiving each treatment. A lack of overlap in the true distribution of pretreatment characteristics between treatment groups implies a violation of this assumption. Exchangeability, also called ignorable treatment assignment assumption, assumes that all true confounders are measured and observed. This condition assumes that the set of observed pretreatment covariates is sufficiently rich, and it includes all variables directly influencing both treatment and outcome.[23] However, it is not possible to test this assumption. In this study, this assumption is violated, because we could not include morbid and co-morbid factors due to the inherent nature of the data.

Causal-Estimands:

Average Treatment Effect (ATE) and Average Treatment Effect Among the Treated (ATT) are two types of causal effect estimands that can be used to derive a summary measure of treatment effect. The definition of ATE and ATT are slightly different, and the choice of this estimand depends on the research question and the target population under treatment. ATE is a better estimand if every member in the study population has equal opportunity to receive treatment. However, ATT is the choice of estimand if the research question of interest is to study the effectiveness of one treatment and if it were replacing an alternate form of treatment. Clearly, in our analysis, ATE was the choice of estimand since it is plausible that every SAH patient had an equal chance of receiving one of the three vasopressors, and the research question focuses on the relative effect of three vasopressors on the outcome, mortality. Moreover, in a binary treatment setting, ATE of a treatment (*T*) relative to a control (*C*) is the difference in the mean outcome, had the study population was administrated *T* versus *C*. Likewise, in a setting with 3 treatments, there

exist three pairwise ATEs, and they can be interpreted as the difference in mean outcomes had the study population was administrated one treatment k vs. a different treatment $\grave{k}$. ATE can be calculated as below[26]

$$ATE_{k,\grave{k}} = E\left(Y(k) - Y\left(\grave{k}\right)\right) = E(Y(k)) - E(Y\left(\grave{k}\right))$$

After defining the casual estimands appropriate for the research question, we can adjust for the estimated propensity scores in one of the four ways: matching, stratifying, weighting and covariate adjustment. In our analysis, we applied weighting via Inverse Probability Treatment Weights (IPTW) of subjects to create a synthetic sample in which treatment assignment is independent of the measured pretreatment covariates.[27] The idea behind IPTW is to up-weight those who are underrepresented and down-weight those who are overrepresented.[23] The weight of a treated subject is defined as the inverse of its propensity score ($w_i = 1/e(X_i)$, and the weight of a control subject is defined as the inverse of one minus its propensity score ($w_i = 1/(1 - e(X_i))$).

PS estimation for multiple treatments:

Imbens[34] extends Rosenbaum and Rubin's work to multiple treatment cases by introducing the concept of the generalized propensity score. The generalized propensity score r (*t*, *X*) is the conditional probability of receiving treatment *t* given the pre-treatment variables *X*,

$$r(t,X) = P(T_i = t|\, X_i) = E\{I(t)|X = X\}$$

where *T* denotes the treatment received and takes on a value in a set *T*, and *I(t)* is the indicator of receiving treatment *t*. Unbiased estimation of average treatment effects conditioning on the generalized propensity score requires the assumption of weak ignorability. Weak ignorability is defined as pairwise independence of the treatment assignment and potential outcomes.[35] If weak

ignorability holds, the inverse of the generalized propensity scores at each treatment level can be used as weights to estimate the mean potential outcome, $\widehat{\mu_t}$ for each treatment level as shown below

$$\widehat{\mu_t} = \frac{\sum_{i=1}^{n} T_i[t] Y_i w_i[t]}{\sum_{i=1}^{n} T_i[t] w_i[t]}$$

Where $w_i = 1/P(T_i = t|\, X_i)$

Using GBM to obtain PS:

PS for multiple treatments can be estimated using multinomial regression or machine learning methods. Considering the high dimensional nature of the SAH data, we used a machine learning approach based on trees called the generalized boosted model (GBM). GBM can typically obtain robust propensity score weights with better balance properties than the multinomial logistic method.[23] GBM estimation involves an iterative process with multiple regression trees to capture complex and nonlinear relationships between treatment assignment and the pretreatment covariates without over-fitting the data.[23] GBM is fit to separate pairs of treatments which consider one treatment as a holdout and estimates $P(I_i(t)|X)$, where $I_i(t) = \{1 \text{ if } T_i = t, 0 \text{ otherwise}\}$ using the estimated odds ratios of the probability assigned to each treatment versus the probability assigned to the holdout treatment. Depending on the choice of holdout treatment, the estimated probabilities might differ since this method relies on the binary estimation of the subsamples of the population. The estimated propensity scores from each of the GBM fits is used to compute the weights. McCaffrey states that the general intuition behind this approach is that, for any given treatment, $T_i = t$, estimating weights for a treatment group only requires knowing the probability

that each case assigned to this group received that assignment rather than one to any other treatment[23].

We also used Toolkit for Weighting and Analysis of Nonequivalent Groups (twang 1.3) packages in R for evaluation of propensity scores and associated weights. The multinomial propensity scores (mnps) function in this package can handle more than two treatment conditions, and it relies on GBM to generate numerous regression trees for estimating the propensity score model, which could lead to the best balance between the treated and reference groups.[23] In mnps, the GBM is fitted to separate pairs of treatment, not to a multinomial model such as that in a multinomial logistic regression model. Then, the estimated propensity scores from each of the GBM fits are used to compute the weights. GBM uses stopping rules to select optimal iterations to yield the best balance. These are based on summary statistics (mean, max) of the absolute standardized mean difference (es) or the Kolmogorov Smirnov (ks) statistic. For covariate k (k = 1, … K) and treatment t (t = 1, … M), the population standardized mean difference, PSB_{tk} is given by[23]

$$PSB_{tk} = \left|\overline{X_{kt}} - \overline{X_{kp}}\right| \Big/ \widehat{\sigma_{kp}}$$

Where $\overline{X_{kt}}$ is the propensity score weighted mean of the covariate, $\overline{X_{kp}}$ and $\widehat{\sigma_{kp}}$ the unweighted mean and standard deviation of the covariate for the pooled sample across all treatments. This statistic measures the similarity of each treatment group to the population in terms of covariate means both before and after weighting. Population KS (PKS) statistic for assessing balance with multiple treatments is given by[23]

$$PKS_{tk} = \sup_{x} \left|EDF_{tk}(x) - EDF_{pk}(x)\right|$$

Where $EDF_{tk}(x) = \sum_{i=1}^{n} I(X_{ik} \leq x)/n$ is the unweighted empirical distribution function for the pooled sample across all treatments.

Assessing Covariate Balance:

After calculating IPTW, it is important to assess the balance. Treatment groups are considered balanced if the absolute standardized mean differences are less than about 0.2. The "plot ()" function was used to assess the balance graphically.

GLM approach for subgroup analysis

The first step in the GLM approach for subgroup analysis was using a lasso to identify the covariates that were associated with treatment response. Glmnet package was used to implement logistic regression modeling with the lasso penalty. The binary outcome variable, mortality, was coded as 1 (death at hospital discharge or discharge to hospice), and 0 (survived), and the treatment is coded as 1 (=Dopamine), 2 (=Phenylephrine), 3(=Norepinephrine). The covariates include demographic factors (age, race, gender and marital status) and pretreatment medications. The base model obtained after lasso penalized feature selection will be of the form below

$$\Pr(Y = 1|X) = \log\left(\frac{\mu}{1-\mu}\right) = \beta_0 + \beta_1 Dopamine + \beta_2 Norepinephrine + \beta_3 X_{c1} + \ldots\ldots + \beta_k X_{ck}$$

where $X_{c1} \ldots . X_{ck}$ represent selected covariates using lasso penalized feature selection.

In order to identify the differential effect of treatment among the selected covariates, the treatment-covariates interaction effect for each covariate is included in the base model. For selected covariate X_{c1}, the interaction model will of the form below

$$\Pr(Y = 1|X) = \beta_0 + \beta_1 Dopamine + \beta_2 Norepinephrine + \beta_3 X_{c1} + \ldots\ldots + \beta_k X_{ck} + \beta_{k+1} Dopamine \times X_{c1} + \beta_{K+2} Norepinephrine \times X_{c1}$$

Similarly, the treatment-covariates interaction effect was studied for every selected covariate. The base model and interaction model were then compared using a likelihood ratio test. To correct for multiple comparisons, we used the false discovery rate (FDR) approach.

Machine learning approach for subgroup analysis using MOB

Model-based Recursive Partitioning (MOB) is a tree-based recursive partitioning method proposed by Zeileis et al.[28] MOB combines both regression modeling and tree structure for automated detection of the patient subgroup. This involves fitting an M-type parametric model at terminal nodes of the tree structure. Like other methods based on classification and regression trees (CART), MOB also uses recursive partitioning methods to create a decision tree. CART finds optimal cuts or dichotomizations in variables, and then finds the best cut among the groups created by previous cuts, hence identifying higher-order covariate interactions.[29] MOB is more suitable for subgroup analysis since it restricts interactions to treatment × covariate interactions.

MOB algorithm works by fitting a logistic regression model with mortality as the response variable and vasopressor treatment as a control variable, and it fits on all observations using maximum likelihood estimation. For n observation of Y_i , $(i = 1, \ldots n)$ the parametric model can be fitted, and its parameter estimate $\hat{\theta}$ can be calculated by minimizing objective function $\Psi(Y, \theta)$ such that

$$\hat{\theta} = \underset{\theta}{argmin} \sum_{i=1}^{n} \Psi(Y_i, \theta_i)$$

Further, the algorithm tries to detect pretreatment covariate effects by testing parameter instability. The parameter instability was assessed with respect to each pretreatment covariate while controlling for the vasopressor treatment. The parameter instability was assessed to determine which variable should be used for partitioning. This is assessed using a score function, $\widehat{\Psi} = \Psi(Y_i, \hat{\theta})$, for systematic deviations from their mean 0 over possible covariates. These deviations are captured by the empirical fluctuation of the process[21]

$$W_j(t) = \hat{J}^{-1/2}\, n^{-1/2} \sum_{i=1}^{\lfloor nt \rfloor} \widehat{\Psi}\sigma(X_{ij}) \qquad (0 \le t \le 1)$$

where $\sigma(x_{ij})$ is the ordering permutation which gives the antirank of the observation X_{ij} in the vector of covariates X_j, and $\hat{J}$ is the covariance matrix $Cov\left(\Psi(Y_i, \hat{\theta})\right)$. This framework of testing parameter stability is established by Zeileis et al.[30]

If the overall null hypothesis of no instability (for any of the parameters) is not rejected, we assume there are no (further) subgroups.[29] However, if the null hypothesis is rejected, this implies there is instability. Then, the variable with the smallest p-value will have the highest parameter instability. Further, the sample is partitioned into subgroups with respect to this split variable using a binary split. When more than two splits are possible, the split which minimizes the objective function of the model in the two resulting subgroups will be chosen.[29] Then the new model will be fitted in each subgroup, and the parameters of each model will be tested again for instability. This process is repeated until no more parameter instability is observed, or minimum sample is reached. MOB uses Bonferroni correction to adjust for multiple testing. We used 'partykit' package and glmtree function in R to implement model-based Recursive Partitioning.

Sensitivity Analysis adjusting for heart rate and blood pressure

Since only 39% of the non-traumatic SAH patient's heart rate and blood pressure values were available, a separate analysis on subjects (n= 940) was conducted to adjust for these variables.

Results

Summary Statistics

The baseline characteristics of 2415 non-traumatic SAH patients are displayed in Table 1. Besides demographic and vital sign factors, 750 pretreatment medication variables were also considered in our analysis. Since the medical diagnoses such as heart disease and diabetes were unknown before or after the vasopressor treatment, we assumed that the pretreatment medications would capture the clinical subgroups. Except for gender, most of the baseline characteristics had a statistically significant difference between the treatment groups.

Table 1: Baseline characteristics of non-traumatic SAH patients by vasopressor treatment group

Characteristic N (%)	Total 2415	Phenylephrine 1251 (51.8)	Dopamine 492 (20.3)	Norepinephrine 672 (27.8)	p-value <0.001
Age, years, mean (SD)	56.5 (14.8)	56.28 (14.34)	58.1 (15.4)	55.62 (15.16)	0.021
Male Sex, n (%)	879 (36.4)	459 (36.6)	172 (34.9)	248 (36.9)	0.75
Race, n (%)					<0.001
African American	549 (22.7)	243 (19.4)	114 (23.1)	192 (28.5)	
White	1609 (66.6)	867 (69.3)	325 (66.0)	417 (62.0)	
Other	257 (10.6)	141 (11.2)	53 (10.7)	63 (9.4)	
Marital Status, n (%)					0.004
Divorced	258 (10.6)	133 (10.6)	46 (9.3)	79 (11.7)	

Married	1098 (45.4)	576 (46.0)	225 (45.7)	297 (44.2)	
Single	637 (26.4)	327 (26.1)	113 (22.9)	197 (29.3)	
Widowed	180 (7.4)	79 (6.3)	48 (9.7)	53 (7.8)	
Unknown	242 (10.0)	136 (10.8)	60 (12.2)	46 (6.8)	
Mortality, n (%)	879 (36.4)	317 (25.3)	251 (51.0)	311 (46.2)	<0.001
Blood pressure (mmHg), mean (SD)					
Systolic*	122.6 (28.8)	126.34 (27.17)	115.31 (31.74)	117.44 (29.99)	<0.001
Diastolic*	65.6 (16.2)	66.87 (15.08)	65.41 (17.31)	63.06 (17.7)	<0.001
Heart Rate, mean* (SD)	84.14 (21.76)	80.44 (18.75)	82.94 (23.65)	91.96 (24.63)	<0.001

** calculated for subsample n=940*

Multinomial lasso and GBM

The first objective in our analysis was to create covariate balanced data. For this purpose, we conducted variable selection using L1-penalized generalized linear models prior to GBM. We used the cv.glmnet function in R with 10-fold cross validation to estimate an optimal value of λ. The prediction error curve as a function of λ is shown in Figure 1. We chose the $\lambda = 0.012$ that yielded the lowest cross-validated prediction error. After the initial multinomial lasso variable selection for the treatment model, 159 out of 750 medication covariates were selected along with paitents' age, race and marital status. We included all these selected factors into GBM to generate a propensity score model and the corresponding IPTWs. The 'mnps' function in the twang package was used with parameter settings of "n.trees" = 6000, "estimand" = ATE and "stop.method"= "es.mean", "ks.mean". Figure 2 shows the graphical assessment of balance achieved through inverse probability treatment weighting. Figure 2a shows that maximum absolute standardized mean difference (AMSD) decreases for all pretreatment covariates after weighting for both

stopping rules. As a rule of thumb, AMSD less than 0.2 is required to achieve a good balance. In our case, after weighting, AMSD less than 0.3 is achieved, which could be considered as a moderate to good balance.[26] The Figure 2b shows that pairwise minimum p-values increased after propensity score weighting (hollow circle). The 45-degree line in this figure shows qline corresponds to the quantiles in the uniform distribution. When the null hypothesis is true, the p-values from the independent test should follow a uniform distribution.[31] The large deviation of the solid circle from this 45-degree line shows the lack of balance before weighting. After weighting, the fact that the p-value has been brought closer to the 45-degree line confirms there was an improvement in covariate balance between the treatment groups.

Figure 1: Plot of the misclassification error and the number of variables in the model as functions of -log(λ) for the 10-fold cross-validation analyses. The lowest point in the curve indicates the optimal lambda.

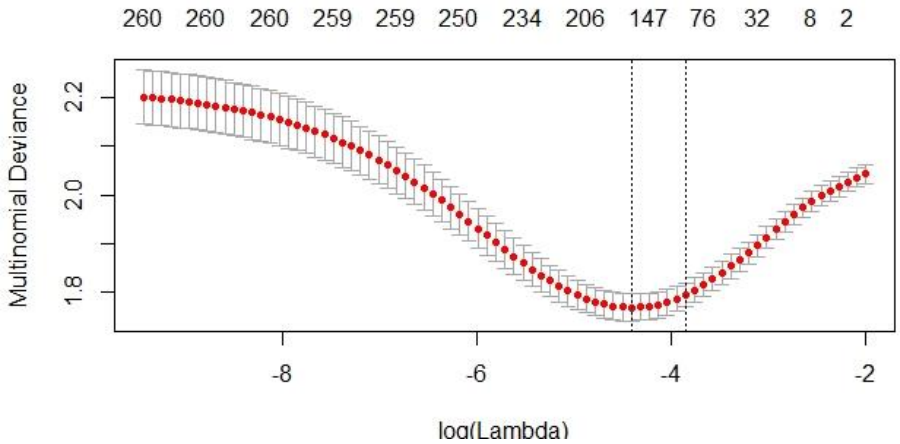


Figure 2a: The maximum pairwise absolute standardized mean differences of the pretreatment covariates before and after weighting based on es.mean and ks.mean stopping rule.

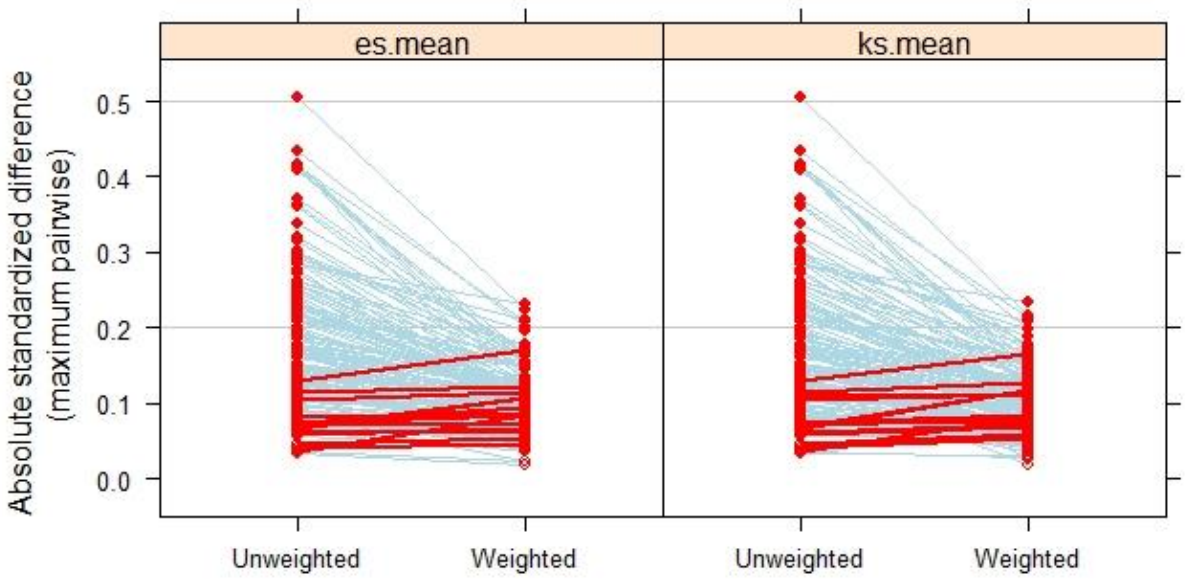
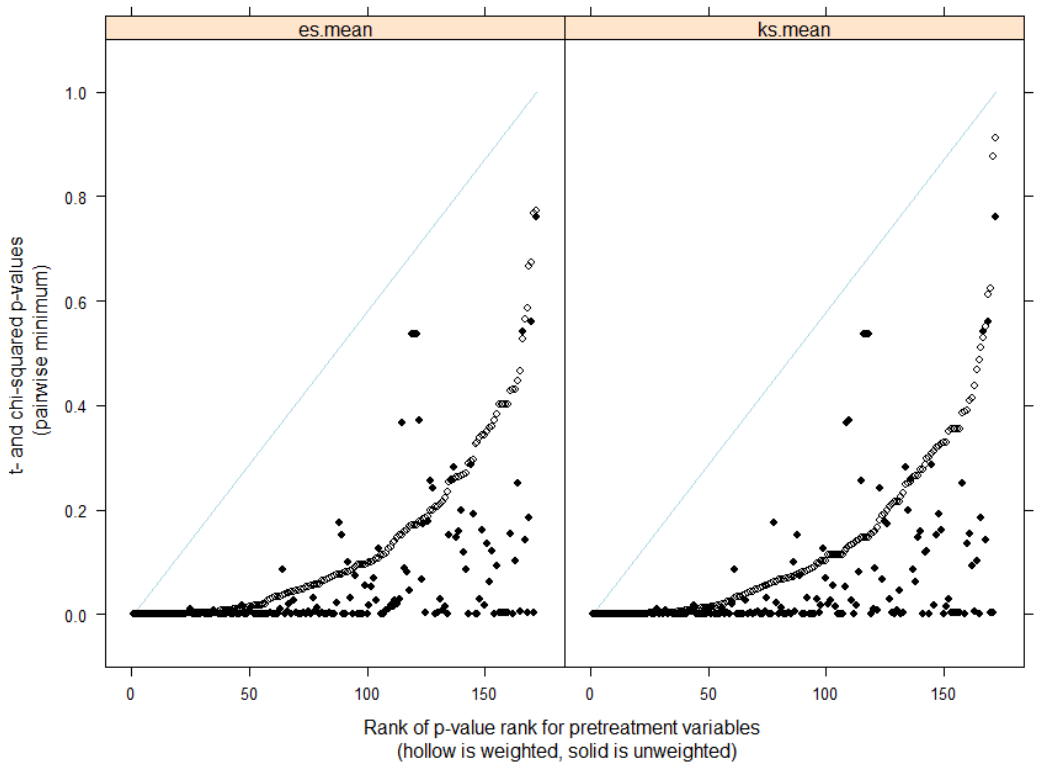


Figure 2b: QQ plot comparing the quantiles of the observed p-values to the quantiles of the uniform distribution



Subgroup Analysis: Binary lasso and Interaction studies

After constructing balanced data, we proceeded with subgroup analysis. At first, we conducted variable selection using L1-penalized generalized linear models. However, here we were only interested in finding covariates associated with mortality. So, we used cv.glmnet with 10-fold crossvalidation and estimated an optimal value of λ. We observed that age, total fluids administered, and vasopressor treatment were associated with mortality. Additionally, 58 out of the 750 pretreatment medications were associated with mortality. Utilizing these significant factors, the base model was constructed as below:

$$\Pr(Mortality = 1|X) = \beta_0 + \beta_1 Dopamine + \beta_2 Norepinephrine + \beta_3 X_{c1} + \ldots\ldots + \beta_{63} X_{c60}$$

where $X_{c1} \ldots\ldots X_{c60}$ represent selected covariates using lasso penalized feature selection. Further, with respect to each selected 60 covariates, 60 interaction models were constructed as follows:

Interaction model 1:

$$\begin{aligned}\Pr(Mortality = 1|X) &= \beta_0 + \beta_1 Dopamine + \beta_2 Norepinephrine + \beta_3 X_{c1} + \ldots\ldots + \beta_{63} X_{c60} \\ &+ \beta_{64} Dopamine \times X_{c1} + \beta_{65} Norepinephrine \times X_{c1}\end{aligned}$$

Interaction model 60:

$$\begin{aligned}\Pr(Mortality = 1|X) &= \beta_0 + \beta_1 Dopamine + \beta_2 Norepinephrine + \beta_3 X_{c1} + \ldots\ldots + \beta_{63} X_{c60} \\ &+ \beta_{64} Dopamine \times X_{c60} + \beta_{65} Norepinephrine \times X_{c60}\end{aligned}$$

We used svyglm function in R to incorporate with IPTW obtained from GBM. Further, a likelihood ratio test was used to compare the base model with each interaction model. In order to account for multiple comparisons, we used the FDR correction procedure. The results indicated that only pretreatment medication of amitriptyline (*Med_lable_34563*) had significant vasopressor-medication interaction. Table 2 shows the coefficient estimate and the p-value for the significant interaction model.

Further, in order to understand the vasopressor effect in the presence and absence of amitriptyline, we calculated various effects as shown in Table 3. Utilizing these effects, we calculated the odds ratio and the corresponding confidence interval. For instance, the odds ratio comparing dopamine to phenylephrine among those who had amitriptyline was calculated by subtracting effect (4) – effect (2). Similarly, the odds ratio comparing norepinephrine to phenylephrine among those who had amitriptyline was calculated by subtracting effect (6) – effect (2). Likewise, effect (3) –effect (1) gave the odds ratio comparing dopamine to phenylephrine among those who did not have amitriptyline. Besides, to find the odds ratio comparing norepinephrine to phenylephrine among those who did not have amitriptyline, we subtracted the effect (1) from effect (5). The results are shown as a forest plot in Figure 3.

Table 2: Coefficient estimate (p-value) for outcome model fitted with IPTW, vasopressor treatment indicator (with ref: phenylephrine) and its interaction term with amitriptyline after adjusting for pretreatment variables.

Effects	Model with Vasopressor- amitriptyline interaction
Intercept	0.47 (<0.001)
Dopamine	0.129 (<0.001)
Norepinephrine	0.125 (<0.001)
Amitriptyline	0.269 (<0.001)
Dopamine× Amitriptyline	-0.49(<0.001)
Norepinephrine× Amitriptyline	Not estimated due to small sample sizes

Table 3: Calculation of various treatment effect based on the interaction model.

Effect	Vasopressor	Pretreatment Medication (Prt Med.) (Each variable from $X_{c1}, \ldots, X_{C60}$ with corresponding coefficients $\beta_3, \ldots, \beta_{63}$)	Logit(P(Y=1) where the coefficient for Prt Med is represented as β_i
1	Phenylephrine (Ph)	Not given	β_0
2	Phenylephrine (Ph)	Given	$\beta_0 + \beta_i$
3	Dopamine (Do)	Not given	$\beta_0 + \beta_1$
4	Dopamine (Do)	Given	$\beta_0 + \beta_1 + \beta_i + \beta_{64}$
5	Norepinephrine (Nor)	Not given	$\beta_0 + \beta_2$
6	Norepinephrine (Nor)	Given	$\beta_0 + \beta_2 + \beta_i + \beta_{65}$

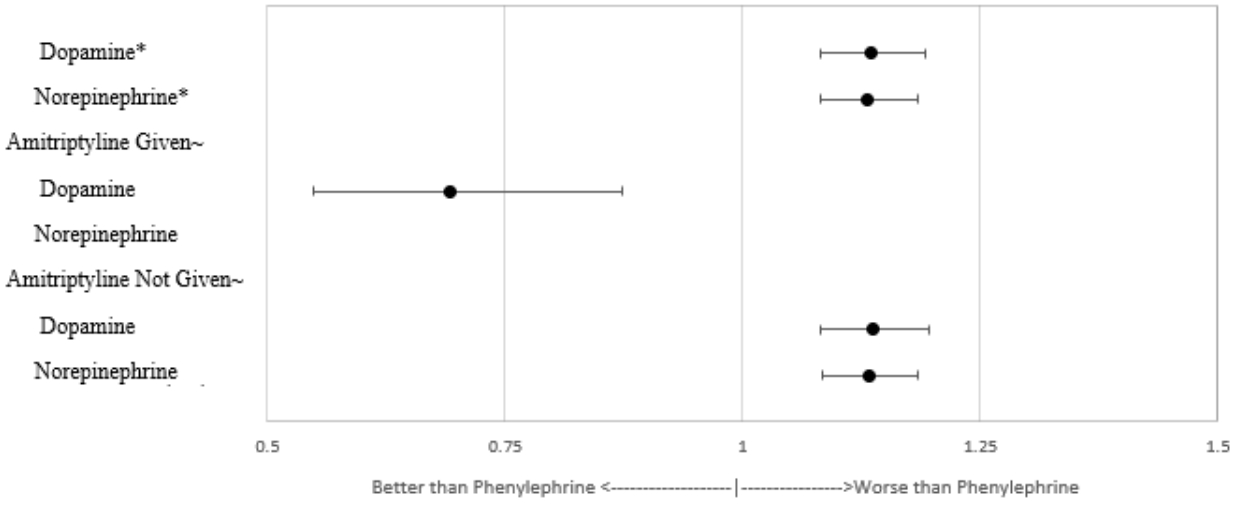
The results showed that dopamine and norepinephrine had higher mortality compared to phenylephrine based on the base model without interaction. The odds ratio of mortality of dopamine compared to that of phenylephrine among those who had amitriptyline was 0.69, and 95% CI was (0.55,0.88); however, the odds ratio among those who did not have amitriptyline was 1.13, and 95% CI was (1.08, 1.19). We did not observe a statistically significant interaction effect between amitriptyline and norepinephrine. When we examined the number of subjects who had amitriptyline stratified by the type of vasopressor and mortality (Table 5), the results showed that amitriptyline was never administered to the subjects who had norepinephrine.

Table 4: The OR and 95% confidence interval for the association of treatment with mortality using IPTW after adjusting for pretreatment variables

Effects	OR	(95% Confidence interval)
For the base model without interaction		
Dopamine	1.13	(1.07, 1.19)
Norepinephrine	1.13	(1.08, 1.18)
Phenylephrine	Ref	Ref
Amitriptyline given		
Dopamine	0.69	(0.55, 0.88)
Norepinephrine	Not estimated due to small sample sizes	
Phenylephrine	Ref	Ref
Amitriptyline not given		
Dopamine	1.13	(1.08, 1.19)

Norepinephrine	1.13	(1.08, 1.18)
Phenylephrine	Ref	Ref

Figure 3: Forest plot showing odds ratios and confidence interval comparing dopamine and norepinephrine to phenylephrine, in presence and absence of amitriptyline.



* Odds ratio for the base model without interaction. ˜ Odds ratio for amitriptyline interaction model. All odds ratio is adjusted for significant pretreatment covariates from the lasso model.

Table 5: The number of subjects who had amitriptyline stratified by type of vasopressor and mortality

Mortality		amitriptyline	
	Vasopressor	Not given	Given
Survived	phenylephrine	926	8
	dopamine	240	1
	norepinephrine	361	0
Dead	phenylephrine	309	8
	dopamine	251	0
	norepinephrine	311	0

Subsample analysis

As described above, subgroup analysis was repeated on complete cases (N=940) for which heart rate and blood pressure values were available. After conducting the propensity score modeling, covariate balance was assessed, and a balance as in shown Figure 2a and 2b was observed. Further, the lasso and interaction effect analysis were conducted. In subsample analysis, none of the

medications was observed to have a significant interaction effect with the vasopressor treatment group.

Model-based Recursive Partitioning (MOB)

We used MOB, a machine learning approach to capture the differential vasopressor treatment effect across the subgroups. In order to implement MOB, we utilized glmtree function in R. To account for covariate imbalance within the data, we used IPTW calculated through GBM. As shown in Figure 4, for sample size N=2,415, ondansetron, fentanyl, mannitol, lidocaine, and age were the partitioning variables selected for the tree. The terminal node in Figure 4 has a boxplot comparing mortality (as y) across the vasopressor treatment group (as x) for that subgroup. The fitted lines are the mean predicted probabilities in each group. Since automatically generated boxplot is unclear, Table 6 is included to show the distribution of the number of subjects for each subgroup stratified by treatment and mortality.

The results showed that among those who had ondansetron, the treatment effect on mortality followed a similar pattern across the age subgroups: phenylephrine was associated with lower odds of mortality, and norepinephrine was associated with higher odds of mortality. Moreover, it can be observed that across all terminal nodes, the rate of mortality was lower among those who had ondansetron compared those who did not have ondansetron. Between all nodes, phenylephrine was consistently observed to be associated with lower mortality, and dopamine seemed to be associated with higher odds of mortality among those who did not have ondansetron.

In Figure 5, an equivalent observation was found for a subsample (n=940). Based on the subsample data consisting of complete cases with heart rate and blood pressure values, the significant subgroup was identified only with the pretreatment medication of ondansetron. The mortality is

much lower among those who had ondansetron where phenylephrine was associated with lower odds of mortality. Among those who had ondansetron, norepinephrine was associated with higher odds of mortality. However, among those who did not have ondansetron, dopamine was associated with higher odds of mortality.

Figure 4: Logistic-regression-based tree for the n= 2,415. The plots in the leaves show boxplot for mortality by vasopressor treatment group

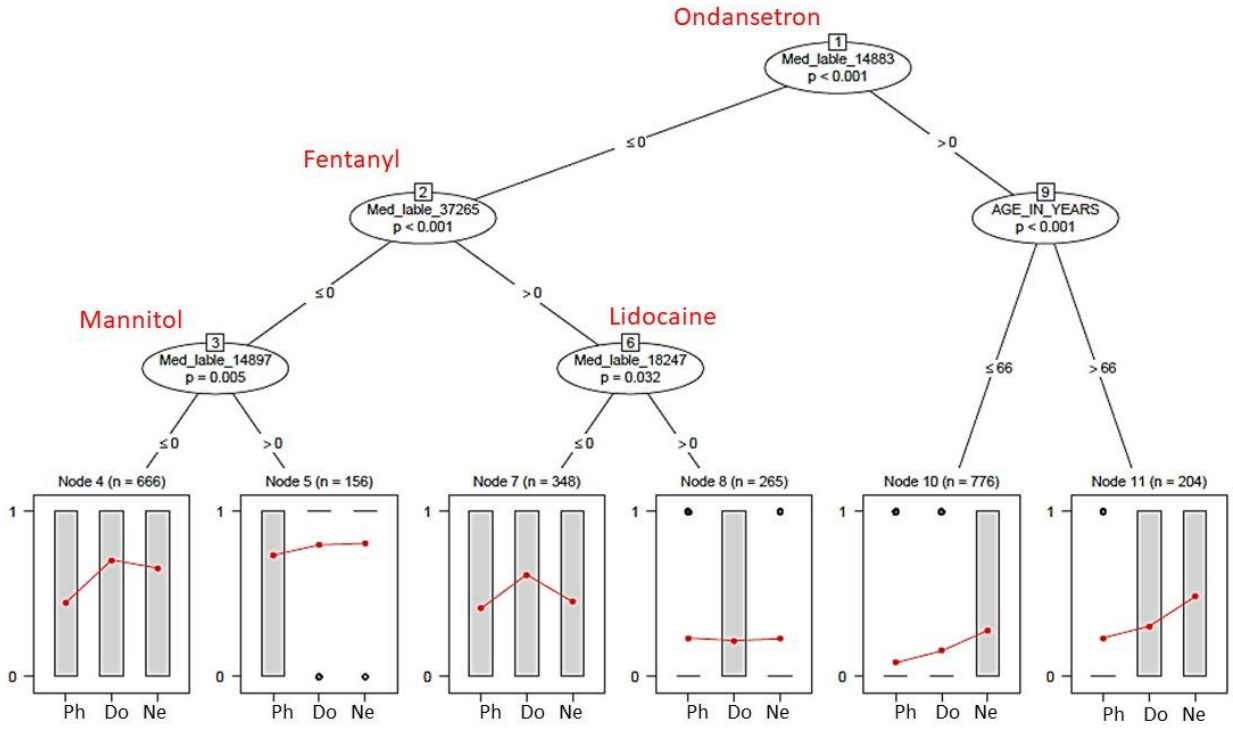
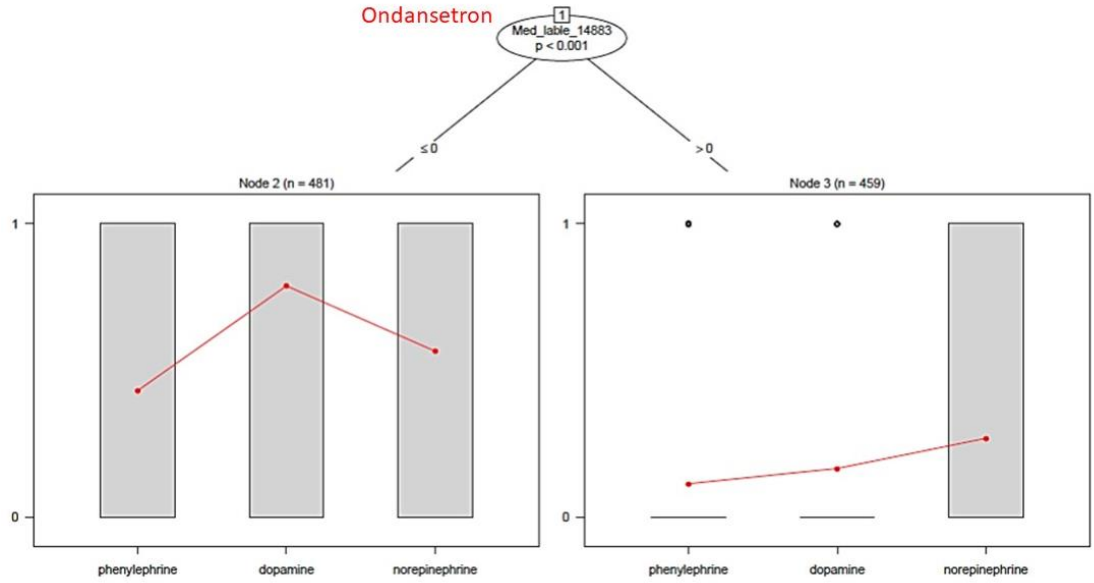


Table 6: Number of subjects for each subgroup stratified by treatment and mortality

Node	Mortality	Phenylephrine	dopamine	norepinephrine	N
4	0	155	61	71	666
	1	97	141	141	
5	0	23	12	5	156
	1	50	38	28	
7	0	112	21	61	348
	1	72	31	51	

8	0	125	30	50	265
	1	36	12	12	
10	0	423	90	152	776
	1	36	19	56	
11	0	96	27	22	204
	1	26	10	23	

Figure 5: Logistic-regression-based tree for the n = 940. The plots in the leaves show boxplots for mortality by vasopressor treatment group



Discussion

A previous study on Cerner-based SAH patient records showed that among the three vasopressors, dopamine, phenylephrine, and norepinephrine, phenylephrine administration has been associated with a lower rate in mortality. On this account, we were interested in investigating if this overall effect was generalizable to all SAH patients. In order to examine this, a subgroup analysis was conducted to study the differential vasopressor treatment effect among the SAH patients. We performed subgroup analysis through a generalized linear model and a machine learning tree-based approach.

Pretreatment administration of ondansetron was a significant factor since it had a differential treatment effect based on the tree-based method. Ondansetron is selective serotonin 5-HT3 (5- hydroxytryptophan 3) antagonist that is used to prevent nausea and vomiting associated with several conditions including anesthesia and surgery.[32] Based on the mob results (Figures 4 and 5), among those who had ondansetron, the predicted probability of mortality is lower compared to those who did not have ondansetron after controlling for the vasopressor treatment group. Besides among those who had ondansetron, phenylephrine was associated with the lowest mortality, and norepinephrine was associated with the highest mortality. Metanalysis on the effects of ondansetron on spinal anesthesia-induced hypotension showed that though ondansetron is commonly used as an antiemetic, it can also reduce the incidence of hypotension and bradycardia as well as the need of vasopressor requirement.[33] Our observation based on mob also aligns with this result: with the administration of ondansetron, norepinephrine and dopamine can be avoided since ondansetron along with phenylephrine has a better survival rate. However, whether vasopressor is required along with ondansetron needs to be further examined.

Based on mob analysis, we also observed that among those who had ondansetron, there were two subgroups by age. The SAH patients aged 66 or older had a higher probability of mortality compared to the younger patients. However, both age subgroups had lower mortality compared to those who did not have ondansetron except in one subgroup. This subgroup included those who did not have ondansetron but had fentanyl and lidocaine. Fentanyl is a synthetic opioid that is 80-100 times stronger than morphine and lidocaine, and it is an anesthetic and antiarrhythmic drug. Though this subgroup falls under those who did not have ondansetron, they had mortality in a similar range to those who had ondansetron with respect to all three vasopressors.

In contrast, amitriptyline was the only medication that was observed to be significant based on a generalized linear model approach. The odds ratio of mortality of dopamine compared to that of phenylephrine among those who had amitriptyline was lower (OR= 0.69 and 95% CI = 0.55,0.88) compared to those who did not have amitriptyline (OR= 1.13 and 95% CI = 1.08, 1.19). However, this medication was administrated to only 17 out of the 2,415 subjects. Hence, the inference based on this result could be spurious. A literature search showed a tricyclic antidepressant agent, such as amitriptyline was often administrated to SAH survivors who had a sign of depression.[36] This may indicate amitriptyline is not a pretreatment medication; however, for amitriptyline toxicity, vasopressor could be used to manage hypotension, even though are not first-line therapy.

On other hand, when fitting a generalized linear model on the subsample for which heart rate and blood pressure values are available, we did not observe any significant interactions. A major difference between the generalized linear model approach and mob-based subgroup analysis was that, in the mob, the multiple-level interactions with treatment were studied. However, in our generalized model approach, we only looked at the first-order interaction since interpreting higher-order interactions are difficult.

This study has several limitations. First, the EHR data are retrospective, and the vasopressor treatments were not administered randomly. However, it would be very hard and expensive to address this limitation in clinical trial settings because SAH is rare. So, this database is an important source to investigate these questions. Second, all the important factors that led to the choice of the right vasopressor may not be captured in the database; therefore, there is a high likelihood of having not included unmeasured confounding effects such as information on clinical severity level and baseline diagnostic and labs. If there are unmeasured confounding effects, the

exchangeability assumption for propensity modeling will be violated. Consequently, there could be a potential for selection bias. Of course, the medications may have captured some of these variables, and we observed moderate to good covariate balance after the inverse probability weighting. However, including any unmeasured confounders in future studies would be helpful to understand differential treatment with respect to morbidities and comorbidities. Third, because each hospital could choose what data elements to upload, the database has a high percentage of missing blood pressure values. In sensitivity analyses; however, we did not detect important differences between those that had and those who did not have blood pressure values; and when repeating the analysis in the subgroup that did have blood pressure, the results were mainly consistent. Another limitation is that, in our analysis primary exposure of interest is the first vasopressor administered to eligible patients, however about 6% patients were given all three vasopressors, and about 30% patients received a second vasopressor after the initial vasopressor. Since we were interested in association of mortality with first choice vasopressor, our analysis did not consider the second and third vasopressor administrated. But it is important acknowledge the fact that second and third medication also influence on vasopressor treatment effect.

Conclusion

In this thesis, we conducted an exploratory analysis to identify subgroups within the SAH patient data based on their vasopressor treatment effect on mortality. The machine learning approach (mob) and statistical approach (GLM) generated different results. This difference, however, could be the result of the inherent differences of the two methodologies. In GLM, we only looked at the first order interaction, and we did not observe any significant interaction effect. Whereas in mob, higher-order interactions were studied, and significant interaction effects were observed with

second-degree and third-degree interactions. The mob result suggests that among those who had ondansetron, norepinephrine was associated with the highest mortality rate, and among those who did not have ondansetron, dopamine was associated with the highest mortality. For the subgroup who did not have ondansetron, but fentanyl and lidocaine, there is no significant difference in the mortality rate between the three vasopressors. Though literature suggests ondansetron can reduce the incidence of hypotension, ondansetron is not widely used in medical practice for this purpose, and hence the clinical significance of the interaction effect of an antiemetic agent such as ondansetron and vasopressor is not clearly understood. It could be considered that the observed significant interaction effect was a marker of certain patients' conditions. But in order to rule out this further analyses considering the pretreatment diagnosis and relevant lab factors are required. However, those who has ondansetron had better survival compared to those who did not have ondansetron with respect to all the three vasopressor. Besides, in general, among three vasopressors, phenylephrine was associated with lower mortality compared to dopamine and norepinephrine.

References

1. Xu J, Murphy SL, Kochanek KD, Bastian B, Arias E. Deaths: Final Data for 2016. Natl Vital Stat Rep 2018; 67:1-76.
2. Rincon F, Rossenwasser RH, Dumont A. The epidemiology of admissions of nontraumatic subarachnoid hemorrhage in the United States. Neurosurgery 2013; 73:217–22; discussion 2–3.
3. Gijn JV. Neurological Disorders Chapter 36 subarachnoid hemorrhage. Elsevier. 2003.
4. Abulhasan YB, Alabdulraheem N, Simoneau G, Angle MR, Teitelbaum J. Mortality after Spontaneous Subarachnoid Hemorrhage: Causality and Validation of a Prediction Model. World Neurosurg. 2018; 112:799-811.
5. de Rooij NK, Linn FH, van der Plas JA, Algra A, Rinkel GJ. Incidence of subarachnoid haemorrhage: a systematic review with emphasis on region, age, gender and time trends. J Neurol Neurosurg Psychiatry. 2007; 78:1365–1372.
6. Welty TE, Horner TG. Pathophysiology and treatment of subarachnoid hemorrhage. Clin. Pharm.1990; 9:35–39.
7. Roy B, McCullough LD, Dhar R, Grady J, Wang YB, Brown RJ. Comparison of initial vasopressors used for delayed cerebral ischemia after aneurysmal subarachnoid hemorrhage.Cerebrovasc Dis. 2017;43:266–271.
8. Lantigua H, Ortega-Gutierrez S, Schmidt JM, Lee K, Badjatia N, Agarwal S, Claassen J, Connolly ES, Mayer SA. Subarachnoid hemorrhage: who dies, and why? Crit Care. 2015; 19:309.
9. Muzevich, K.M., Voils, S.A. Role of vasopressor administration in patients with acute neurologic injury. Neurocrit Care. 2009; 11:112–119.
10. Brookes ST, Whitely E, Egger M, Smith GD, Mulheran PA, Peters T. Subgroup analyses in randomized trials: risks of subgroup-specific analyses; power and sample size for the interaction test. J Clin Epidemiol 2004; 57:229-236.
11. Ferreira1 JC, Patino CM. Subgroup analysis and interaction tests: why they are important and how to avoid common mistakes. J Bras Pneumol. 2017; 43: 162-162.
12. Burke JF, Sussman JB, Kent DM, Hayward RA. Three simple rules to ensure reasonably credible subgroup analyses. BMJ. 2015; 351:h5651.
13. Brookes ST, Whitley E, Peters TJ, Mulheran PA, Egger M, Davey, Smith G. Subgroup analyses in randomised controlled trials: quantifying the risks of false-positives and false-negatives. Health Technol Assess 2001; 5:1-56
14. https://www.himss.org/library/ehr
15. Forrest GN, Schooneveld, TCV, Kullar R, Schulz, LT, Duong, P and Postelnick, M. Use of electronic health records and clinical decision support systems for antimicrobial stewardship. Clin. Infectious Dis. 2014; 59:122–133.

16. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. Journal of the american medical informatics association. JAMIA. 2013; 20:144–51.
17. Weng C, Appelbaum P, Hripcsak G, Kronish I, Busacca L, Davidson KW, Bigger JT. Using EHRs to integrate research with patient care: promises and challenges. J Am Med Inform Assoc. 2012; 19:684–7.
18. Hawkins DM. The problem of overfitting. J Chem Inf Comput Sci 2004; 44:1-12.
19. Zhang Z. Too much covariates in a multivariable model may cause the problem of overfitting. J Thorac Dis 2014; 6: E196-E197.
20. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996; 49:1373-1379.
21. Hosmer DW Lemeshow S. Applied logistic regression. John Wiley & Sons, New York. 2000
22. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Journal of the Royal Statistical Society, Series B 1995; 57: 289–300.
23. McCaffrey DF, Griffin BA, Almirall D, Slaughter ME, Ramchand R, Burgette LF. A tutorial on propensity score estimation for multiple treatments using generalized boosted models. Stat Med. 2013; 32:3388-414.
24. Rosenbaum PR and Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika. 1983; 70:41–55.
25. D'Agostino Jr RB. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group. Stat Med. 1998; 17:2265-2281.
26. Thomas GJ. Utilizing Propensity Score Methods for Ordinal Treatments and Prehospital Trauma Studies. Texas Medical Center Dissertations (via ProQuest). 2017; AAI10681743.
27. Austin PC, Stuart EA. Moving towards best practice when using inverse probability of treatment weighting (IPTW) using the propensity score to estimate causal treatment effects in observational studies. Stat Med. 2015; 356:3661-79.
28. Zeileis A, Hothorn T, Hornik K. Model-Based Recursive Partitioning. Journal of Computational and Graphical Statistics. 2008; 17:492–514.
29. Seibold H., Zeileis A, & Hothorn T. Model-based recursive partitioning for subgroup analyses. International Journal of Biostatistics. 2016; 12,45–63.
30. Zeileis A, Hornik K. Generalized M-Fluctuation Tests for Parameter Instability. Statistica Neerlandica, 2007; 61: 488-508.
31. Greg R, Dan M, Andrew M, Beth AG, Lane B. Toolkit for Weighting and Analysis of Nonequivalent Groups. Version 1.5. CRAN. 2017.
32. Wilde M. & Markham A. Drugs (1996) 52: 773
33. Gao L,Zheng G, Han J, Wanga Y, Zhenga J. Effects of prophylactic ondansetron on spinal anesthesia-induced hypotension: a meta-analysis. Int J Obstet Anesth, 2015; 24:335-343.

34. Imbens GW. The role of the propensity score in estimating doseresponse functions. Biometrika 2000:87, 706 – 710.
35. Stepien, KM Performance of propensity score methods in the presence of heterogeneous treatment effects. UMD Theses and Dissertations. 2016
36. Huttunen J, Lindgren A, Kurki MI, et al. Antidepressant use after aneurysmal subarachnoid hemorrhage: a population-based case-control study. Stroke 2016; 47:2242–2248.

Appendices

Appendix I

Summary of variables

Variables	Type	Number of variables/dummy variables
Vasopressor treatment	Categorical	2
Medication generic names prior to vasopressor treatment	Binary	750
Age	Numeric	1
Race	Categorical	2
Gender	Binary	1
Marital status	Categorical	3
Vital signs (HR, SBP, DBP)	Numeric	3
Total fluids administered	Numeric	1